

A GUIDE TO IMPLEMENTING LLMs IN PRACTICE

Retina specialists can use AI to streamline their communication, if they do it carefully. Here's how.

By Fatima Kalabi, MD; Mohammad Hossein Amirhosseini, BSc, MSc, PhD; and Rodrigo Anguita, MBBS, MD, FEBO, FICO



Large language models (LLMs) are not reading OCT scans for us yet, but they are changing how we talk to patients, draft correspondence, and prepare trainees for oral examinations. Understanding what they can and cannot do is now a practical clinical skill. For a broad overview of what LLMs are and their potential uses in the clinic, see *AI Chatbots in the Retina Clinic*.



Here, we explore *how*, exactly, retina specialists can implement these tools safely in their workflow.



DRAFTING PATIENT EDUCATION

Retinal disease can present significant communication challenges. The difference between wet and dry AMD, the rationale for repeated intravitreal injections, or the urgency of a retinal detachment repair may be obvious to the specialist but bewildering to a newly diagnosed patient. LLMs can produce a first draft of a plain-language explanation, adjust reading level, or reformat content for a patient portal.^{1,2}

In a cross-sectional study of ChatGPT-4 (OpenAI) responses to common questions about retinal detachment, macular hole, and epiretinal membrane surgery, retina specialists judged most answers to be appropriate. However, the authors cautioned against treating these tools as factual authorities and highlighted persistent concerns about credibility and readability in specialized contexts.³

The key phrase is *first draft*. Without careful prompting, models may overstate the likelihood of visual improvement, omit the need for ongoing treatment, or miss return precautions entirely. The clinician is always responsible for the final message.

Patient Education Example Prompt: Draft a [350]-word handout at a sixth-grade reading level for a patient newly diagnosed with [neovascular AMD] who is starting [intravitreal anti-VEGF injections]. Include: what the condition is; that treatment aims to preserve rather than restore vision; that treatment continues alongside monitoring rather than as a fixed course; and a separate list of symptoms needing same-day contact such as increasing pain, worsening redness, sudden reduction in vision, new floaters or flashes. Exclude drug names, treatment intervals, success rates and prognostic figures. Write [CONFIRM] wherever a statement depends on this patient's findings.

Prompt design: Whenever possible, give the model something to write rather than something to avoid. Stating that *treatment aims to preserve rather than restore vision* is more reliable than *do not overstate benefit*. Where a prohibition is unavoidable, make it specific: *no drug names, intervals, or success rates* is quickly checkable, whereas *nothing misleading* is not.

Requiring the red flags be a separate list is a formatting instruction doing safety work. Content requested only by topic tends to be absorbed into the surrounding prose,



where it softens. For example, if you ask for warning symptoms, the LLM may weave them into a closing paragraph of reassurance, phrased as things to raise at the next visit. Instead, ask the LLM to give warning symptoms their own heading or bulleted list to ensure they are a discrete, scannable instruction. This also does clinical work, as a patient scanning a handout reads lists rather than paragraphs. Specify format wherever content must be easy to find, not merely present. The same precision applies to readability. *Sixth-grade reading level* produces consistent output where *simple language* does not.

Finally, the [CONFIRM] placeholder turns silent gap-filling into a visible prompt for the clinician. This is the highest-yield instruction in the set, and one worth reusing across every prompt that follows.

DRAFTING LETTERS

High-volume retina clinics generate large amounts of letters to general practitioners, optometrists, and referring colleagues. A specialist can provide de-identified bullet points and prompt AI to produce a structured letter; specifying diagnosis, visual acuity, key OCT findings, treatments, and management plan. The model is not deciding whether to treat, extend, or observe, but helping to express a decision already made.

Correspondence Example Prompt: Convert the bullet points below into a structured letter to a referring [optometrist]. Use only the information given. Do not add findings, diagnoses, recommendations, or follow-up intervals that are not stated. If a standard letter section cannot be completed from the material provided, write “not stated” rather than filling the gap. Keep to [200] words in a formal clinical tone. [Paste de-identified bullets]

Prompt design: The key instruction here is the gap-handling rule. *Use only the information given* is necessary but weak on its own, because it tells the model what not to do without giving it an alternative action. When a conventional letter section has no input, the model resolves the conflict by completing the pattern. Supplying the fallback output *not stated* gives it a compliant way to leave the gap; it turns an omission you would have to detect into one you can see. Generalize the technique by specifying what to produce whenever you forbid something.

DRAFTING CALL SCRIPTS

LLMs can assist in drafting symptom-based scripts for call handlers, FAQ responses for common queries, and text for clinic websites or appointment reminders. Triage decisions should follow existing clinical protocols; the LLM’s role is to make approved guidance clearer, not to replace it.

Call Script Example Prompt: Below is our triage protocol for calls about [flashes and floaters]. Rewrite it as a call-handler script with the questions in the order they should be asked and the resulting action for each answer. Change no threshold, timeframe, or action. Add no new questions. List anything ambiguous under “for clinical review” rather than resolving it. [Paste protocol]

Prompt design: This is a transformation task rather than a generation task, and it is the safest pattern available. The clinician includes the clinical content in the prompt, and the model restructures the language, reducing the review to comparing the output against the source.

Change no threshold, timeframe, or action names the dimensions along which potential drift occurs, which works better than a general instruction to stay faithful to the original. The ambiguity clause matters for a subtler reason: A model asked to rewrite an unclear instruction will clarify it, improving readability while quietly changing clinical meaning. Routing ambiguity to a separate list preserves it as a question for the clinician rather than resolving it silently. Wherever an approved document already exists, prefer this pattern to drafting from scratch.

STUDY TOOLS

Evidence for LLMs as educational tools has improved substantially as models have advanced. Trainees can use them to generate single best-answer questions, Objective Structured Clinical Examination scenarios, viva prompts, and synthetic retina cases, including consent role-plays for intravitreal injection, vitrectomy, scleral buckle, pneumatic retinopexy, or macular hole surgery. The value lies in structuring knowledge and practicing communication; factual claims should always be checked against trusted sources.

KEY TAKEAWAYS

- ▶ Large language models are text-generation tools that can help retina specialists draft patient education and correspondences, teach cases, and research documents.
- ▶ Without careful prompting, AI models may overstate the likelihood of visual improvement, omit the need for ongoing treatment, or miss return precautions entirely.
- ▶ Safe adoption requires no patient-identifiable data, institution-approved platforms, careful prompting, and clinician review of every output.



Study Example Prompt: Write [one] best-answer question at fellowship level on [the management of rhegmatogenous retinal detachment], each question with four distractors and a short explanation. Do not cite papers, trials, or numerical outcomes. Where management is genuinely contested, say so in the explanation rather than presenting one option as definitive.

'GPT-5.6 Thinking' (OpenAI) response: A 62-year-old pseudophakic patient presents with a macula-involving rhegmatogenous retinal detachment. Examination shows multiple retinal breaks in different quadrants, fixed retinal folds, and a circumferential epiretinal membrane causing retinal shortening. What is the most appropriate primary management?

- A. Pneumatic retinopexy with cryotherapy
- B. Segmental scleral buckle without drainage
- C. Pars plana vitrectomy, membrane peeling, retinopexy, and internal tamponade
- D. Demarcation laser photocoagulation around the detachment
- E. Intravitreal expansile gas followed by delayed laser retinopexy

Answer: C

Explanation: The fixed folds, circumferential membrane, and retinal shortening indicate established proliferative vitreoretinopathy. Successful repair requires removal of vitreoretinal traction, dissection of proliferative membranes, closure of all retinal breaks and internal tamponade. Retinectomy may also be required if the retina remains shortened after membrane removal. Pneumatic retinopexy, laser alone, and an isolated segmental buckle do not adequately address extensive traction or multiple breaks in different quadrants. The choice of gas versus silicone oil, and whether to add an encircling buckle, remains dependent on the detachment configuration, severity of proliferative vitreoretinopathy, patient factors, and surgeon preference.

Prompt design: Exclude by task, not by subject. Models fabricate references and outcome figures readily, but write reliable question stems and plausible distractors, so the prompt avoids the first and asks for the second. The final clause addresses tone rather than accuracy. Explanations read as authoritative whether the evidence supports it or not, so where surgeon techniques differ, AI will present one option as settled unless told otherwise. Specify how the output should read, not only what it should contain.

LIMITATIONS AND MYTHS

LLMs can hallucinate—ie, generate plausible but false statements. In retina, this could mean a wrong dosing interval, an inaccurate trial result, or an invented contraindication. A systematic review of LLMs in

ophthalmology found that most studies reported accuracy but few addressed calibration or uncertainty, with substantial heterogeneity in pooled diagnostic performance.⁴

Hallucination Example Prompt: What is the licensed treatment interval for faricimab (Vabysmo, Genentech/Roche) in diabetic macular edema, and how does it differ from the aflibercept 8 mg (Eylea HD, Regeneron) regimen?

Claude Opus 5 (Anthropic) response, in part: Following 4 monthly loading doses, faricimab may be extended to fixed 16-week dosing in eligible patients, with the YOSEMITE and RHINE protocols permitting extension in 4-week increments based on central subfield thickness and BCVA criteria assessed at weeks 20, 24, and 28.

Hallucination Analysis: The general approach is right, the drug and trials are real, and the mechanism is described with appropriate vocabulary. However, the specifics such as the maximum interval, the increment size, the assessment weeks, and the extension criteria are assembled from the statistical regularities of how anti-VEGF regimens are described; they are not from the protocol or the label. Some of it may be correct, but nothing in the output distinguishes which parts.

Two features make this more dangerous than a fabricated citation. A reference can be checked quickly and its absence is unambiguous. A dosing statement, however, must be checked against a label or protocol by someone who already suspects it may be wrong. In this example, the error is partial rather than total, so the surrounding accurate material lends the invented figures credibility.

Avoiding hallucinations is a matter of task selection rather than prompt wording. No instruction prevents an AI model from producing numbers when numbers are what a question demands. Asking an LLM to flag uncertainty helps only where it is uncertain, and here it is not. The tractable move is to stop asking. Regulatory and dosing facts have authoritative sources, and AI can help interpret those documents, not supply the facts themselves. Where a specific figure is needed, paste the relevant section and ask the LLM to summarize or reformat it, which converts the task from recall, where the model is unreliable, to transformation, where it is reliable.

Data protection deserves particular attention. Removing a patient's name or date of birth does not reliably de-identify clinical notes. Laterality, procedure dates, rare diagnoses, visual acuity history, and unusual treatment sequences can combine to make a patient recognizable, particularly in smaller or highly specialized services. Patient-identifiable or potentially identifiable information should never be entered into public tools.



GETTING STARTED SAFELY

The safest starting point is low-risk, non-identifiable work: generic patient education materials, teaching cases, departmental workflow documents, journal club outlines, and rewriting approved text in simpler language. A practical prompt specifies the audience, task, length, tone, limits, and instructs the model not to invent missing information.

Seven Tips That Make an AI Prompt Work

1. **Task and audience:** clarify who is reading this, and why
2. **Format/length:** specify word count, headings, and tone
3. **Reading level:** name a grade level, not "simple"
4. **Source material:** state "use only what I provide below"
5. **Prohibitions:** state "include no invented findings, figures, intervals, or citations"
6. **Gap handling:** stipulate that the AI "write 'not stated' rather than filling any gaps"
7. **Uncertainty:** ask that it "flag anything I should verify"

Lines 1–3 determine whether the output is usable, and lines 4–7 determine whether it is safe. Every prohibition should be paired with an alternative action, or the model will resolve the gap on your behalf.

Institution-approved or enterprise deployments should be preferred over public consumer tools, particularly for anything approaching clinical documentation. Platform approval does not remove the need for output review.

Enterprise deployment at scale is already arriving. In June 2026, NHS England announced that Microsoft 365 Copilot would be made available to more than 500,000 clinicians and support staff, following a trial across 90 organizations in which administrative support saved an average of 43 minutes per user per day.⁶ Tools embedded in an institutional environment can offer clearer data-governance arrangements than public consumer products, but the same principles apply.

Departments benefit from an internal policy defining appropriate tasks, approved tools, what information must not be entered, how outputs should be reviewed, and whether AI assistance requires disclosure in academic or patient-facing materials.

TOWARD MULTIMODAL RETINA ASSISTANTS

Retina is a high-stakes specialty; small errors in timing, interpretation, or wording can affect vision. Until larger multimodal AI assistants are validated and regulated, LLMs are best understood as textual power tools in today's clinic. They can explain, summarize, draft, and teach. They can also hallucinate, omit, and sound confident when wrong. Used with curiosity and caution, LLMs may become a useful conversational layer around the AI already reshaping, rather than replacing, the human expertise at the center of the specialty. ■

AI disclosure: The authors used Claude Opus 5 (Anthropic) in the preparation of this manuscript for language refinement of specific sections. All content was reviewed, verified, and edited by the author, who takes full responsibility for the accuracy and integrity of the work.

1. Anguita R, Makuloluwa A, Hind J, Wickham L. Large language models in vitreoretinal surgery. *Eye (Lond)*. 2024;38(4):809-810. doi.org/10.1038/s41433-023-02751-1
2. Ferro Desideri L, Roth J, Zinkemagel M, Anguita R. Application and accuracy of artificial intelligence-derived large language models in patients with age-related macular degeneration. *Int J Retina Vitreous*. 2023;9(1):71. doi.org/10.1186/s40942-023-00511-7
3. Momenaei B, Wakabayashi T, Shahlaee A, et al. Appropriateness and readability of ChatGPT-4-generated responses for surgical treatment of retinal diseases. *Ophthalmol Retina*. 2023;7(10):862-868. doi.org/10.1016/j.oret.2023.05.022
4. Zhang Z, Zhang H, Pan Z, et al. Evaluating large language models in ophthalmology: systematic review. *J Med Internet Res*. 2025;27:e76947. doi.org/10.2196/76947
5. Mihalache A, Huang RS, Patil NS, et al. Chatbot and Academy Preferred Practice Pattern guidelines on retinal diseases. *Ophthalmol Retina*. 2024;8(7):723-725. doi.org/10.1016/j.oret.2024.03.013
6. NHS England. 500,000 NHS staff to get new artificial intelligence tools to help free up more time for patients. Published June 2026. Accessed June 25, 2026. tinyurl.com/5xuj2r84

FATIMA KALABI, MD

- Ophthalmology FY2 Resident Doctor, Barking, Havering and Redbridge University Hospitals NHS Trust, London, United Kingdom
- Financial disclosure: None

MOHAMMAD HOSSEIN AMIRHOSSEINI, BSC, MSC, PHD

- Associate Professor in Computer Science and Digital Technologies, University of East London, London, United Kingdom
- Financial disclosure: None acknowledged

RODRIGO ANGUITA, MBBS, MD, FEBO, FICO

- Vitreoretinal Consultant, Moorfields Eye Hospital National Health System Foundation Trust, London, United Kingdom
- Associate Professor, Institute of Ophthalmology, University College London, London, United Kingdom
- rodrigoanguita@gmail.com
- Financial disclosure: None

Moorfields Advanced Vitreoretinal Congress (MAVERICK)

A new international meeting dedicated to advancing excellence in vitreoretinal surgery

January 29 - 30, 2027

One Great George Street, Westminster, London



Moorfields
Education

MAVERICK 2027
Moorfields Advanced Vitreoretinal Congress

